

# LYAPUNOV EXPONENTS OF LINEAR COCYCLES OVER MARKOV SHIFTS

ELAÍAS C. MALHEIRO AND MARCELO VIANA

ABSTRACT. The Lyapunov exponents of  $GL(2)$ -cocycles over Markov shifts depend continuously on the underlying data, that is, on the matrix coefficients and the Markov measure transition probabilities.

## 1. INTRODUCTION

The notion of Lyapunov exponent goes back to the late 19th century work of A. M. Lyapunov [12] on the stability of solutions of differential equations: the so-called first method of Lyapunov amounts to saying that, under suitable conditions, one has exponential stability whenever the exponents are all negative.

It turns out that these numbers encode very important information on the behavior of the dynamical system. A striking illustration of this fact is the theory of (non-uniformly) hyperbolic systems, initiated by Oseledets [14] and Pesin [15, 16]. In a few words, it asserts that smooth diffeomorphisms and flows whose Lyapunov exponents do not vanish exhibit a very rich geometric structure, including invariant stable and unstable laminations that are absolutely continuous (see for instance [5, Appendix C]), from which a refined description of the dynamics can be drawn.

In this work we are particularly concerned with the dependence of Lyapunov exponents on the underlying system. In addition to its intrinsic interest, this problem is also connected to the issue of when do Lyapunov exponents vanish. For example Bourgain, Jitomirskaya [7, 6] proved that the Lyapunov exponent of certain Schrödinger cocycles with analytic potential vary continuously with the energy parameter and used that fact to obtain an explicit lower bound for the Lyapunov exponent. From this they deduced that such cocycles have Anderson localization.

While there is a number of other situations where Lyapunov exponents are known to vary continuously, and even analytically, under perturbations of the system (see [17, Section 10.6] and the references therein), it is also known that in general their behavior is rather wild. Again, an especially striking illustration is the discovery, due to Mañé [13], that in the realm of  $C^1$  area-preserving surface diffeomorphisms the only continuity points for the Lyapunov exponents are the Anosov maps (which exist only on the torus) and those maps whose exponents vanish almost every point. A complete proof was given by Bochi [1] and this has also been extended to arbitrary dimension, by Bochi, Viana [3, 2].

The derivative of a  $C^1$  diffeomorphism is a  $C^0$  linear cocycle, of course. For cocycles that are better than just continuous, the picture seems to be richer and is far from being understood. In this direction, the second author conjectured around

---

*Date:* October 7, 2014.

The authors are supported by CNPq, FAPERJ and PRONEX-Dynamical Systems.

2004 that the Lyapunov exponents vary continuously restricted to the subset of fiber-bunched Hölder-continuous  $GL(2)$ -cocycles. For definitions and more precise statements, see [5, Section 12.4] and [17, Section 10.6]. The fiber-bunching condition can not be omitted, as observed in [17, Section 9.3]. But continuity should hold within any family of cocycles admitting invariant holonomies.

The first result was obtained in 2009 by Bocker, Viana [4], who proved continuity for locally constant cocycles over a Bernoulli shift. It is this result that we now extend to the Markov case: *The Lyapunov exponents of  $GL(2)$ -cocycles over Markov shifts depend continuously on the matrix coefficients and the transition probabilities.* The complete statement is given in Theorem A below.

Our method is very different from the one in [4] and may be viewed as an extension of the method used in [17, Chapter 10] to give an alternative proof of the theorem of Bocker-Viana. In either case, the starting point is the notion of stationary measure and the fact that the stationary measures completely determine the Lyapunov exponents (see Furstenberg [8] and Ledrappier [11]). Then, both arguments are based on estimating the stationary measures of nearby cocycles, to prove that they can not accumulate too much on the neighborhood of a single point. This, they achieve in very different ways.

Now, the notion of stationary measure does not really make sense outside of the Bernoulli case. In the present Markov setting, we are able to substitute it with the notion of *stationary measure vector*, which we introduce here. While this is a fairly straightforward extension of the notion of stationary measure, stationary measure vectors are a lot more difficult to handle because, in a few words, we need the different vector components to be quite homogeneous. Indeed, a good part of our efforts in Sections 8 through 9 is devoted to establishing such homogeneity.

This notion of stationary measure vector has a natural counterpart in the general setting of the conjecture mentioned previously, and the methods we develop here should be useful for proving that conjecture. But the analysis of stationary measure vectors in such generality has yet to be carried out in full.

This work is organized as follows. In Section 2 we give the precise context and statement of our main result. In Section 3 we define stationary measure vector and prove a few useful properties. In Sections 4 and 5 we adapt to our setting an argument of Furstenberg, Kifer [10] showing that any possible discontinuity point must exhibit an invariant subspace that is a kind of repeller for the action of the cocycle in projective space. Our task is then to show that continuity holds even in the presence of such a repeller. In Section 6 we reduce this task to proving a key statement, Theorem 6.1. In Section 7 we recall the notion of energy of a measure, which has a central role in our arguments. Then, in Sections 8 and 9, we use it to prove Theorem 6.1, thus completing the proof.

## 2. STATEMENT OF MAIN RESULT

Let  $X = \{1, \dots, q\}$  and  $f : M \rightarrow M$  be the shift map on  $M = X^{\mathbb{N}}$ . We use  $x = (i_n)_n$  to represent a generic element of  $M$ . For any  $m \geq 0$ ,  $n \geq 1$  and  $j_0, \dots, j_{n-1} \in X$ , denote

$$[m; j_0, \dots, j_{n-1}] = \{x \in M : i_{m+s} = j_s \text{ for } 0 \leq s \leq n-1\}.$$

Let  $P = (P_{i,j})_{i,j \in X}$  be a stochastic matrix and assume that  $P$  is *aperiodic*, that is, there exists  $N \geq 1$  such that  $P_{i,j}^N > 0$  for all  $i, j \in X$ . By Perron-Fröbenius,

there exists a unique vector  $p = (p_i)_{i \in X}$  such that

$$p_i > 0 \quad \text{and} \quad \sum_{s \in X} p_s = 1 \quad \text{and} \quad \sum_{s \in X} p_s P_{s,j} = p_j \quad \text{for every } i, j \in X.$$

Let  $\mu$  be the associated Markov measure on  $M$ : given any  $m \geq 0$  and  $n \geq 1$ ,

$$\mu([m; j_0, \dots, j_{n-1}]) = p_{j_0} P_{j_0, j_1} \cdots P_{j_{n-2}, j_{n-1}} \quad \text{for every } j_0, \dots, j_{n-1} \in X.$$

Then  $\mu$  is invariant under  $f$  and the system  $(f, \mu)$  is mixing.

We use  $u, v, w$  to represent the lines spanned by vectors  $\underline{u}, \underline{v}, \underline{w} \in \mathbb{R}^d \setminus \{0\}$ . Let  $d \geq 2$  and  $A : X \rightarrow \text{GL}(d)$ . Consider the linear cocycle

$$F : M \times \mathbb{R}^d \rightarrow M \times \mathbb{R}^d, \quad (x, \underline{v}) \mapsto (f(x), A(i_0)\underline{v})$$

and the associated projective cocycle

$$\mathbb{P}(F) : M \times \mathbb{P}(\mathbb{R}^d) \rightarrow M \times \mathbb{P}(\mathbb{R}^d), \quad (x, v) \mapsto (f(x), A(i_0)v).$$

Here and in what follows we use the same notation for an element of  $\text{GL}(d)$  and for its action in projective space.

Denote  $A^n(x) = A(i_{n-1}) \cdots A(i_1)A(i_0)$ , for each  $x \in M$  and  $n \geq 1$ . By Furstenberg, Kesten [9], there exist numbers  $\lambda_+(A, P) \geq \lambda_-(A, P)$  such that

$$\begin{aligned} \lambda_+(A, P) &= \lim_n \frac{1}{n} \log \|A^n(x)\| \\ \lambda_-(A, P) &= \lim_n \frac{1}{n} \log \|A^n(x)^{-1}\|^{-1} \end{aligned} \quad \text{for } \mu\text{-almost every } x \in M.$$

They are called the *extremal Lyapunov exponents* of  $F$  relative to  $\mu$ . We are going to prove:

**Theorem A.** *For  $d = 2$ , the functions  $(A, P) \mapsto \lambda_{\pm}(A, P)$  are continuous, relative to the natural (coefficient-induced) topology in the space of pairs  $(A, P)$ .*

The special case of this theorem when  $\mu$  is a Bernoulli measure, that is, when  $P_{i,j} = p_j$  for every  $i, j \in X$ , is contained in the main result of Bocker, Viana [4]. Recently, Avila, Eskin, Viana announced that, still in the Bernoulli case, the theorem holds in arbitrary dimension  $d \geq 2$ . Although their paper has not yet appeared, the 2-dimensional version of their approach has been presented in [17, Chapter 10] and that was the starting point for our analysis of the Markov case.

It is well known (see [17, Section 9.1]) that the functions  $\lambda_+$  and  $\lambda_-$  are, respectively, upper semi-continuous and lower semi-continuous. Thus, continuity holds automatically at any point such that  $\lambda_-(A, P) = \lambda_+(A, P)$ . In what follows we assume that  $\lambda_-(A, P) < \lambda_+(A, P)$ . We are going to prove continuity for  $\lambda_+$ : to deduce the corresponding statement for  $\lambda_-$  just use Lemma 4.1 below.

When the stochastic matrix  $P$  is just *irreducible* (for every  $i, j \in X$  there exists  $N \geq 1$  such that  $P_{i,j}^N > 0$ ), our arguments remain valid to show that  $A \mapsto \lambda_{\pm}(A, P)$  are continuous. Moreover, one recovers continuous dependence on both variables, as in the conclusion of the theorem, under the additional assumption that the probability vector  $p$  is chosen depending continuously on the matrix in a neighborhood of  $P$  (this is automatic in the aperiodic case).

Most steps in the proof of Theorem A are valid for arbitrary dimension  $d$ . We specialize to  $d = 2$  only when necessary.

## 3. MEASURE VECTORS

The following simple notions have a central part in our arguments. We consider vectors  $\eta = (\eta_i)_{i \in X}$  where each  $\eta_i$  is a (positive) measure on  $\mathbb{P}(\mathbb{R}^d)$ . With any such vector, we associate the *skew-product measure*  $m = \mu \ltimes \eta$  defined on  $M \times \mathbb{P}(\mathbb{R}^d)$  by

$$(1) \quad m = \sum_{i \in X} (\mu \mid [0; i]) \times \eta_i.$$

We call  $\eta$  a *unit vector* if every  $\eta_i$  is a probability measure on  $\mathbb{P}(\mathbb{R}^d)$ . Then  $m$  is a probability measure on  $M \times \mathbb{P}(\mathbb{R}^d)$ .

Let  $\mathcal{P}$  be the operator defined in the space of measure vectors  $\eta$  by

$$(2) \quad (\mathcal{P}\eta)_j(D) = \sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \eta_i(A(i)^{-1}(D))$$

for any  $j \in X$  and any measurable  $D \subset \mathbb{P}(\mathbb{R}^d)$ . We say that  $\eta$  is *P-stationary* if it is a fixed point for  $\mathcal{P}$ , that is, if it satisfies

$$\sum_{i \in X} p_i P_{i,j} \eta_i(A(i)^{-1}(D)) = p_j \eta_j(D)$$

for every  $j \in X$  and every measurable  $D \subset \mathbb{P}(\mathbb{R}^d)$ . In the Bernoulli case, this means that  $\sum_{i \in X} p_i \eta_i(A(i)^{-1}(D)) = \eta_j(D)$  for every  $j$  and every  $D$ . Then the  $\eta_j$  are all equal, and one recovers the more usual notion of stationary measure.

**Proposition 3.1.** *A unit vector  $\eta$  is P-stationary if and only if the probability measure  $m = \mu \ltimes \eta$  is  $\mathbb{P}(F)$ -invariant.*

*Proof.* For any  $j, j_2, \dots, j_n \in X$  and any measurable set  $D \subset \mathbb{P}(\mathbb{R}^d)$ ,

$$\mathbb{P}(F)^{-1}([0; j, j_2, \dots, j_n] \times D) = \bigcup_{i \in X} [0; i, j, j_2, \dots, j_n] \times A(i)^{-1}(D).$$

Thus, using the definition (1),

$$\mathbb{P}(F)_* m([0; j, j_2, \dots, j_n] \times D) = \sum_{i \in X} \mu([0; i, j, j_2, \dots, j_n]) \eta_i(A(i)^{-1}(D)).$$

The right-hand side may be written as

$$\sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \mu([0; j, j_2, \dots, j_n]) \eta_i(A(i)^{-1}(D)) = \mu([0; j, j_2, \dots, j_n]) (\mathcal{P}\eta)_j(D).$$

This proves that  $\mathbb{P}(F)_*(\mu \ltimes \eta) = \mu \ltimes \mathcal{P}\eta$ . The claim is a consequence.  $\square$

The set of unit vectors  $\eta = (\eta_i)_{i \in X}$  has a natural topology induced by the weak\* topology in the space of probability measures on  $\mathbb{P}(\mathbb{R}^d)$ .

**Proposition 3.2.** *The set of P-stationary unit vectors  $\eta$  is non-empty, compact and convex.*

*Proof.* First, note that the operator  $\mathcal{P}$  is continuous. Indeed, if  $(\eta_n)_n$  is a sequence of unit vectors converging to some  $\eta$  then, given any continuous function  $\phi : \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ ,

$$\int \phi d(\mathcal{P}\eta_n)_j = \sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \int (\phi \circ A(i)) d\eta_{n,i}$$

for every  $j \in X$  and every  $n \geq 1$ . Since  $\eta_{n,i} \rightarrow \eta_i$  in the weak\* topology, the expression on the right-hand side converges to

$$\sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \int (\phi \circ A(i)) d\eta_i = \int \phi d(\mathcal{P}\eta)_j.$$

This proves that  $(\mathcal{P}\eta_n)_j$  converges to  $(\mathcal{P}\eta)_j$  in the weak\* topology, for every  $j \in X$ , as claimed.

Next, let  $\xi$  be an arbitrary unit vector and define, for  $j \in X$  and  $n \geq 1$ ,

$$\eta_{n,j} = \frac{1}{n} \sum_{l=0}^{n-1} (\mathcal{P}^l \xi)_j.$$

Let  $\eta$  be an accumulation vector of the sequence  $(\eta_n)_n$ , where  $\eta_n = (\eta_{n,j})_{j \in X}$ . Observe that

$$(\mathcal{P}\eta_n)_j = \frac{1}{n} \sum_{l=0}^{n-1} (\mathcal{P}^{l+1} \xi)_j = \eta_{n,j} + \frac{1}{n} (\mathcal{P}^n \xi)_j - \frac{1}{n} \xi_j.$$

So, given any continuous function  $\phi : M \rightarrow \mathbb{R}$ ,

$$\lim_n \int \phi d(\mathcal{P}\eta_n)_j = \lim_n \left( \int \phi d\eta_{n,j} + \frac{1}{n} \int \phi d(\mathcal{P}^n \xi)_j - \frac{1}{n} \int \phi d\xi_j \right) = \int \phi d\eta_j.$$

This means that  $\eta_j = \lim_n (\mathcal{P}\eta_n)_j$ . We also have  $\lim_n (\mathcal{P}\eta_n)_j = (\mathcal{P}\eta)_j$ , by the continuity of the operator  $\mathcal{P}$ . It follows that  $(\mathcal{P}\eta)_j = \eta_j$  for every  $j \in X$ , which means that  $\eta$  is  $P$ -stationary. This proves that stationary vectors do exist.

Let  $(\eta_n)_n$  be a sequence of  $P$ -stationary unit vectors converging to  $\eta$ . For each  $j \in X$  and  $n \geq 1$ , we have  $(\mathcal{P}\eta_n)_j = \eta_{n,j}$  and  $\lim_n \eta_{n,j} = \eta_j$ . As we have already seen,  $\lim_n (\mathcal{P}\eta_n)_j = (\mathcal{P}\eta)_j$ . Then  $(\mathcal{P}\eta)_j = \eta_j$ , which means that  $\eta$  is a  $P$ -stationary unit vector. This implies the set of  $P$ -stationary unit vectors is closed and, consequently, is compact.

Let  $\eta_1$  and  $\eta_2$  be  $P$ -stationary unit vectors and consider  $\eta = a_1 \eta_1 + a_2 \eta_2$ . For each  $j \in X$ , we have

$$\begin{aligned} \eta_j(D) &= a_1 \eta_{1,j}(D) + a_2 \eta_{2,j}(D) = a_1 (\mathcal{P}\eta_1)_j(D) + a_2 (\mathcal{P}\eta_2)_j(D) \\ &= \sum_{i \in X} a_1 \frac{p_i P_{i,j}}{p_j} \eta_{1,i}(A(i)^{-1}(D)) + \sum_{i \in X} a_2 \frac{p_i P_{i,j}}{p_j} \eta_{2,i}(A(i)^{-1}(D)) \\ &= \sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \eta_i(A(i)^{-1}(D)) = (\mathcal{P}\eta)_j(D). \end{aligned}$$

This proves that the set of  $P$ -stationary unit vectors is convex.  $\square$

**Proposition 3.3.** *The set  $\{(P, \eta) : \eta \text{ is a } P\text{-stationary unit vector}\}$  is closed.*

*Proof.* Let  $(P_k, \eta_k)_{k \geq 0}$  be a sequence converging to  $(P, \eta)$ , where each  $\eta_k$  is a  $P_k$ -stationary unit vector. This means that, for any  $j \in X$  and  $k \geq 1$

$$\eta_{k,j} = (\mathcal{P}_k \eta_k)_j = \sum_{i \in X} \frac{p_{k,i} P_{k,i,j}}{p_{k,j}} A_k(i) * \eta_{k,i}$$

Then, given any continuous function  $\phi : \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ ,

$$\int \phi d\eta_{k,j} = \sum_{i \in X} \frac{p_{k,i} P_{k,i,j}}{p_{k,j}} \int (\phi \circ A_k(i)) d\eta_{k,i}.$$

The left-hand side converges to  $\int \phi d\eta_j$  and the right-hand side converges to

$$\sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \int (\phi \circ A(i)) d\eta_i = \int \phi d(\mathcal{P}\eta)_j,$$

because  $p_{k,i}$  and  $P_{k,i,j}$  and  $\eta_{k,j}$  and  $A_k(i)$  converge, respectively, to  $p_i$  and  $P_{i,j}$  and  $\eta_j$  and  $A(i)$ . Thus,  $\eta_j = (\mathcal{P}\eta)_j$  for every  $j \in X$ . This means that  $\eta$  is a  $P$ -stationary unit vector.  $\square$

A  $P$ -stationary vector  $\eta$  is *atomic* if there exists  $v \in \mathbb{P}(\mathbb{R}^d)$  such that  $\{v\}$  has positive  $\eta_i$ -measure for some  $i \in X$ . Then  $v$  is an *atom* of  $\eta$ .

**Proposition 3.4.** *If  $\eta$  is an atomic  $P$ -stationary unit vector then there exists a finite non-empty set  $\mathcal{L} \subset \mathbb{P}(\mathbb{R}^d)$  such that  $A(i)v \in \mathcal{L}$  and  $\eta_i(\{v\}) > 0$  for every  $v \in \mathcal{L}$  and every  $i \in X$ .*

*Proof.* Let  $\delta$  be the largest  $\eta_i$ -measure of any point  $v \in \mathbb{P}(\mathbb{R}^d)$  for any  $i \in X$ . By assumption,  $\delta > 0$ . Let  $\mathcal{L}$  be the set of all  $w \in \mathbb{P}(\mathbb{R}^d)$  such that  $\eta_j(\{w\}) = \delta$  for some  $j \in X$ . Notice that  $\mathcal{L}$  is non-empty and  $\#\mathcal{L} \leq \#X/\delta < \infty$ , because the  $\eta_i$  are probability measures. Since

$$\eta_j(\{w\}) = \sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \eta_i(\{A(i)^{-1}w\}) \quad \text{and} \quad \sum_{i \in X} \frac{p_i P_{i,j}}{p_j} = 1,$$

we have that

$$(3) \quad \eta_j(\{w\}) = \delta \quad \Rightarrow \quad \eta_i(\{A(i)^{-1}w\}) = \delta \quad \text{for every } i \in X.$$

In particular,  $A(i)^{-1}(\mathcal{L}) \subset \mathcal{L}$  for every  $i \in X$ . Since  $\mathcal{L}$  is finite, this means that  $A(i)^{-1}(\mathcal{L}) = \mathcal{L}$  for every  $i \in X$ . So, given any  $v \in \mathcal{L}$  and  $i \in X$ , we can find  $w \in \mathcal{L}$  such that  $v = A(i)^{-1}w$ . By (3), it follows that  $\eta_i(\{v\}) = \delta$ .  $\square$

Consider the continuous function  $\Phi : M \times \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$  defined by

$$(4) \quad \Phi(x, v) = \log \frac{\|A(x)\underline{v}\|}{\|\underline{v}\|} = \log \frac{\|A(i_0)\underline{v}\|}{\|\underline{v}\|}.$$

Clearly,

$$\int \Phi d(\mu \times \eta) = \sum_{i \in X} p_i \int \log \frac{\|A(i)\underline{v}\|}{\|\underline{v}\|} d\eta_i(v).$$

Results closely related to the next proposition are discussed in [17, Chapter 6], mostly in the Bernoulli context. The proposition is true in any dimension, but we choose to restrict ourselves to the case  $d = 2$ , which is simpler and suffices for all our purposes.

**Proposition 3.5.**

$$\lambda_+(A, P) = \max \left\{ \int \Phi d(\mu \times \eta) : \eta \text{ is a } P\text{-stationary unit vector} \right\}.$$

*Proof.* Let  $m$  be an arbitrary  $\mathbb{P}(F)$ -invariant probability measure on  $M \times \mathbb{P}(\mathbb{R}^2)$ . For every  $(x, v)$  and every  $n \geq 1$ ,

$$\frac{1}{n} \sum_{j=0}^{n-1} \Phi(\mathbb{P}(F)^j(x, v)) = \frac{1}{n} \log \frac{\|A^n(x)\underline{v}\|}{\|\underline{v}\|} \leq \frac{1}{n} \log \|A^n(x)\|.$$

Let  $\tilde{\Phi}(x, v)$  denote the limit of the left-hand side. This is well-defined  $m$ -almost everywhere, by the ergodic theorem, and the previous inequality implies that  $\tilde{\Phi}(x, v) \leq \lambda_+(A, P)$ . So,

$$(5) \quad \int \Phi dm = \int \tilde{\Phi} dm \leq \lambda_+(A, P).$$

In particular,  $\lambda_+(A, P)$  is an upper bound for the integral of  $\Phi$  relative to any  $\mathbb{P}(F)$ -invariant measure of the form  $m = \mu \times \eta$ . We are left to prove that the equality in (5) does hold for some such measure.

It is convenient to consider the natural extension of the cocycle  $F$

$$\hat{F} : \hat{M} \times \mathbb{R}^2 \rightarrow \hat{M} \times \mathbb{R}^2, \quad \hat{F}(\hat{x}, \underline{v}) = (\hat{f}(\hat{x}), A(i_0)\underline{v}),$$

where  $\hat{f} : \hat{M} \rightarrow \hat{M}$  is the shift on  $\hat{M} = X^{\mathbb{Z}}$ . Denote by  $\hat{\mu}$  the  $\hat{f}$ -invariant Markov measure defined on  $\hat{M}$  by the stochastic matrix  $P$ . Clearly,  $(\hat{F}, \hat{\mu})$  has the same Lyapunov exponents as  $(F, \mu)$ . Since we assume that  $\lambda_-(A, P) < \lambda_+(A, P)$ , the Oseledets decomposition of  $(\hat{F}, \hat{\mu})$  has the form

$$\mathbb{R}^2 = E_x^u \oplus E_x^s, \quad \text{with } \dim E_x^u = \dim E_x^s = 1$$

(the invariant bundles  $E^u$  and  $E^s$  are associated with the Lyapunov exponents  $\lambda_+(A, P)$  and  $\lambda_-(A, P)$ , respectively) for  $\hat{\mu}$ -almost every  $\hat{x} = (i_n)_{n \in \mathbb{Z}} \in \hat{M}$ .

Let  $\mathbb{P}(\hat{F}) : \hat{M} \times \mathbb{R}^2 \rightarrow \hat{M} \times \mathbb{R}^2$  be the projective cocycle associated with  $\hat{F}$ . Let  $\pi : \hat{M} \rightarrow M$  and  $\Pi : \hat{M} \times \mathbb{P}(\mathbb{R}^2) \rightarrow M \times \mathbb{P}(\mathbb{R}^2)$ ,  $\Pi(\hat{x}, v) = (\pi(\hat{x}), v)$  and  $\pi_1 : M \times \mathbb{P}(\mathbb{R}^2) \rightarrow M$  and  $\hat{\pi}_1 : \hat{M} \times \mathbb{P}(\mathbb{R}^2) \rightarrow \hat{M}$  be the canonical projections. If  $\hat{m}$  is a  $\mathbb{P}(\hat{F})$ -invariant measure projecting down to  $\hat{\mu}$  (under  $\hat{\pi}_1$ ) then  $m = \Pi_* \hat{m}$  is an  $\mathbb{P}(F)$ -invariant measure projecting down to  $\mu$  (under  $\pi_1$ ): that is an immediate consequence of the relations  $\mathbb{P}(F) \circ \Pi = \Pi \circ \mathbb{P}(\hat{F})$  and  $\pi_1 \circ \Pi = \pi \circ \hat{\pi}_1$ .

Take  $\hat{m}$  to be the probability measure on  $\hat{M} \times \mathbb{P}(\mathbb{R}^2)$  defined by

$$\hat{m}(C \times D) = \hat{\mu}(\{\hat{x} \in C : E_{\hat{x}}^u \in D\}),$$

for any measurable sets  $C \subset \hat{M}$  and  $D \subset \mathbb{P}(\mathbb{R}^2)$ . In other words,  $\hat{m}$  projects down to  $\hat{\mu}$  and its conditional probabilities along the fibers are given by the Dirac masses at the unstable subspace  $E_{\hat{x}}^u$ . It is clear that  $\hat{m}$  is  $\mathbb{P}(\hat{F})$ -invariant, since the measure  $\hat{\mu}$  is  $\hat{f}$ -invariant and the unstable bundle  $E^u$  is  $\hat{F}$ -invariant. So,  $m$  is a  $\mathbb{P}(F)$ -invariant probability measure.

Next, we prove that  $m$  is a skew-product measure. The key observation is that, since the unstable subspace  $E_{\hat{x}}^u$  is determined by the negative iterates of the cocycle alone, each the conditional probability

$$\hat{m}_{\hat{x}} = \delta_{E_{\hat{x}}^u} \text{ depends only on the negative component } x^- = (i_n)_{n < 0} \text{ of } \hat{x} \in \hat{M}.$$

Moreover, the Markov measure  $\hat{\mu}$  is a product measure restricted to each cylinder  $[0; i] \subset \hat{M}$ : there are probability measures  $\mu_i^-$  and  $\mu_i^+$  on  $M^- = X^{\{n < 0\}}$  and  $M^+ = X^{\{n > 0\}}$ , respectively, such that

$$\hat{\mu} \mid [0; i] = p_i(\mu_i^- \times \mu_i^+).$$

Given  $i \in X$  and measurable sets  $C^- \subset M^-$ ,  $C^+ \subset M^+$  and  $D \subset \mathbb{P}(\mathbb{R}^2)$ ,

$$\begin{aligned} \hat{m}(C^- \times \{i\} \times C^+ \times D) &= \int_{C^- \times \{i\} \times C^+} \hat{m}_{\hat{x}}(D) d\hat{\mu}(\hat{x}) \\ &= p_i \mu_i^+(C^+) \int_{C^-} \hat{m}_{x^-}(D) d\mu_i^-(x^-) \end{aligned}$$

whereas

$$\hat{\mu}(C^- \times \{i\} \times C^+) = p_i \mu_i^-(C^-) \mu_i^+(C^+).$$

In particular, the quotient between these two numbers is independent of  $C^+$ . Any subset  $C$  of the one-sided cylinder  $[0; i] \subset M$  may be written as  $C = \{i\} \times C^+$  with  $C^+ \subset M^+$ . Then,

$$\eta_i(D) = \frac{m(C \times D)}{\mu(C)} = \frac{\hat{m}(M^- \times \{i\} \times C^+ \times D)}{\hat{\mu}(M^- \times \{i\} \times C^+)}$$

is independent of  $C$  and defines a probability measure  $\eta_i$  on  $\mathbb{P}(\mathbb{R}^2)$ . Let  $\eta$  be the measure vector formed by these probabilities. For any measurable  $C \subset M$  and  $D \subset \mathbb{P}(\mathbb{R}^2)$ ,

$$m(C \times D) = \sum_{i \in X} m((C \cap [0; i]) \times D) = \sum_{i \in X} \mu(C \cap [0; i]) \times \eta_i(D).$$

This proves that  $m = \mu \times \eta$ .

Finally, consider the function  $\Psi : \hat{M} \times \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$  defined by

$$\Psi(\hat{x}, v) = \log \frac{\|A(i_0)\underline{v}\|}{\|\underline{v}\|} = \Phi(\pi(\hat{x}), v).$$

Clearly,  $\int \Phi dm = \int \Psi d\hat{m} = \int \tilde{\Psi} d\hat{m}$ , where  $\tilde{\Psi}$  is the Birkhoff time average:

$$\tilde{\Psi}(\hat{x}, v) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \Psi(\mathbb{P}(\hat{F})^j(\hat{x}, v)) = \lim_n \frac{1}{n} \log \frac{\|A^n(x)\underline{v}\|}{\|\underline{v}\|}.$$

Since the conditional probabilities of  $\hat{m}$  are concentrated on the Oseledets subspaces  $E_{\hat{x}}^u$ , we have  $\tilde{\Psi}(\hat{x}, v) = \lambda_+(A, P)$  for  $\hat{m}$ -almost every  $(\hat{x}, v)$ . It follows that  $\int \Phi dm = \lambda_+(A, P)$ .

This completes the proof of Proposition 3.5.  $\square$

#### 4. INVARIANT SUBSPACES

Consider any sequence  $(A_k, P_k)_k$  converging to  $(A, P)$ . Let  $p_k$  be the probability vector associated with each  $P_k$ . Moreover,

**Lemma 4.1.**  $\lambda_+(A_k, P_k) + \lambda_-(A_k, P_k)$  converges to  $\lambda_+(A, P) + \lambda_-(A, P)$ .

*Proof.* By the theorem of Oseledets [14] and the ergodicity of  $(f, \mu)$ ,

$$\lambda_+(A, P) + \lambda_-(A, P) = \lim_n \frac{1}{n} \log |\det A^n(x)| = \int \log |\det A| d\mu$$

for  $\mu$ -almost every  $x$ , and analogously for  $\lambda_+(A_k, P_k) + \lambda_-(A_k, P_k)$  for every  $k$ . Now observe that

$$\int \log |\det A_k| d\mu_k = \sum_{i \in X} p_{k,i} \log |\det A_k(i)|$$



converges to

$$\sum_{i \in X} p_i \log |\det A(i)| = \int \log |\det A| d\mu$$

when  $k \rightarrow \infty$ .  $\square$

For each  $k$ , let  $\Phi_k : M \times \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$  be defined as in (4), with  $A_k$  in the place of  $A$ . Moreover, let  $\eta_k = (\eta_{k,i})_{i \in X}$  be a  $P_k$ -stationary unit vector that realizes the largest Lyapunov exponent for  $(A_k, P_k)$ :

$$\lambda_+(A_k, P_k) = \int \Phi_k d(\mu_k \times \eta_k) = \sum_{i \in X} p_{k,i} \int \log \frac{\|A_k(i)\underline{v}\|}{\|\underline{v}\|} d\eta_{k,i}(v).$$

Up to restricting to a subsequence, we may suppose that  $(\eta_k)_k$  converges to some  $\eta$  in the weak\* topology. By Proposition 3.3, the vector  $\eta$  is  $P$ -stationary. Moreover,

$$\int \Phi_k d(\mu_k \times \eta_k) \text{ converges to } \int \Phi d(\mu \times \eta).$$

If  $\eta$  realizes  $\lambda_+(A, P)$  then we are done. In all that follows we suppose that

$$(6) \quad \int \Phi d(\mu \times \eta) < \lambda_+(A, P).$$

**Proposition 4.2.** *Under assumption (6), there exists  $L \in \mathbb{P}(\mathbb{R}^2)$  such that*

$$(7) \quad \eta_i(\{L\}) > 0 \quad \text{and} \quad A(i)L = L \quad \text{for every } i \in X$$

and, for  $\mu$ -almost all  $x \in M$ ,

$$(8) \quad \lim_n \frac{1}{n} \log \|A^n(x)\underline{v}\| = \begin{cases} \lambda_-(A, P) & \text{if } \underline{v} \in L \setminus \{0\} \\ \lambda_+(A, P) & \text{if } \underline{v} \in \mathbb{R}^2 \setminus L. \end{cases}$$

*Proof.* By Proposition 3.1, the skew-product  $m = \mu \times \eta$  is  $\mathbb{P}(F)$ -invariant. So, we may use the ergodic theorem to conclude that

$$\tilde{\Phi}(x, v) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \Phi(\mathbb{P}(F)^j(x, v)) = \lim_n \frac{1}{n} \log \|A^n(x)\underline{v}\|$$

exists for  $m$ -almost every point, is constant on the orbits of  $\mathbb{P}(F)$  and satisfies  $\int \tilde{\Phi} dm = \int \Phi dm$ . Thus, (6) implies that  $\tilde{\Phi} < \lambda_+(A, P)$  on some subset with positive measure for  $m$ .

**Lemma 4.3.** *For  $\mu$ -almost every  $x \in M$  there exists a unique  $L_x \in \mathbb{P}(\mathbb{R}^2)$  such that  $\tilde{\Phi}(x, L_x) < \lambda_+(A, P)$ . Moreover,*

$$(9) \quad L_{f(x)} = A(x)L_x \quad \text{for } \mu\text{-almost every } x \in M.$$

*Proof.* Consider  $\mathcal{Z} = \{(x, v) \in M \times \mathbb{P}(\mathbb{R}^2) : \tilde{\Phi}(x, v) < \lambda_+(A, P)\}$ . This is a measurable,  $\mathbb{P}(F)$ -invariant set with positive measure for  $m$ . Hence, the projection  $Z = \{x \in M : (x, v) \in \mathcal{Z} \text{ for some } v\}$  is a measurable  $f$ -invariant set with positive measure for  $\mu$ . Since  $(f, \mu)$  is ergodic, it follows that  $\mu(Z) = 1$ . This proves existence. Next, given any  $x \in M$ , suppose that there exist two distinct points  $v_1$  and  $v_2$  in projective space such that  $\tilde{\Phi}(x, v_i) < \lambda_+(A, P)$  for  $i = 1, 2$ . Since every vector in  $\mathbb{R}^2$  may be written as linear combination of  $\underline{v}_1$  and  $\underline{v}_2$ , it follows that

$$\lim_n \frac{1}{n} \log \|A^n(x)\| < \lambda_+(A, P).$$

By the definition of  $\lambda_+(A, P)$ , this can only happen on a subset with zero measure for  $\mu$ . That proves uniqueness. Property (9) follows, since the function  $\tilde{\Phi}$  is constant on the orbits of  $\mathbb{P}(F)$ .  $\square$

**Lemma 4.4.** *There is  $L \in \mathbb{P}(\mathbb{R}^2)$  such that  $L_x = L$  for  $\mu$ -almost every  $x \in M$ .*

*Proof.* Let  $Z_*$  be the union over  $i \in X$  of the set of all  $x \in [0; i] \cap Z$  such that  $L_x$  is an atom of  $\eta_i$ . Observe that  $\mu(Z_*) > 0$ , because

$$\sum_{i \in X} \int_{[0; i]} \eta_i(\{L_x\}) d\mu(x) = m(\{(x, L_x) : x \in M\}) = m(\mathcal{Z}) > 0.$$

Given any  $x \in Z_*$ , consider  $i, j \in X$  such that  $x \in [0; i, j]$ . If  $P_{i,j} = 0$  then  $\mu([0; i, j]) = 0$ . Otherwise, using (9),

$$p_j \eta_j(\{L_{f(x)}\}) = \sum_{s \in X} p_s P_{s,j} \eta_s(\{A(s)^{-1} L_{f(x)}\}) \geq p_i P_{i,j} \eta_i(\{L_x\}) > 0,$$

and so  $f(x) \in Z_*$ . This proves that  $Z_*$  is  $f$ -invariant up to measure zero. By the ergodicity of  $(f, \mu)$ , it follows that  $\mu(Z_*) = 1$ .

Since the  $\eta_i$  are finite measures, they have at most countably many atoms. In particular, there exists  $a \in \mathbb{P}(\mathbb{R}^d)$  such that  $Z_a = \{x \in Z_* : L_x = a\}$  has positive  $\mu$ -measure. Then, for any  $k \geq 1$ , we may find  $n_k$  and  $j_{k,0}, \dots, j_{k,n_k}$  such that

$$\frac{\mu(Z_a \cap [0; j_{k,0}, \dots, j_{k,n_k}])}{\mu([0; j_{k,0}, \dots, j_{k,n_k}])} \geq 1 - \frac{1}{k}.$$

Let  $b_k = A(j_{k,n_k-1}) \cdots A(j_{k,0})a$ . Then, using (9) once more,

$$\begin{aligned} \frac{\mu(Z_{b_k} \cap [0; j_{k,n_k}])}{\mu([0; j_{k,n_k}])} &\geq \frac{\mu(f^{n_k}(Z_a \cap [0; j_{k,0}, \dots, j_{k,n_k}]))}{\mu(f^{n_k}([0; j_{k,0}, \dots, j_{k,n_k}]))} \\ &= \frac{\mu(Z_a \cap [0; j_{k,0}, \dots, j_{k,n_k}])}{\mu([0; j_{k,0}, \dots, j_{k,n_k}])} \geq 1 - \frac{1}{k}. \end{aligned}$$

In particular, for  $k \geq 2$ ,

$$\mu(Z_{b_k}) \geq (1 - k^{-1}) p_{j_{k,n_k}} \geq \min_{i \in X} \frac{p_i}{2} > 0.$$

Clearly, there are only finitely many values of  $b_k$  for which this inequality can hold. Then, since  $j_{k,n_k}$  also takes values in a finite set  $X$ , we may choose a subsequence  $k_l \rightarrow \infty$  for which the values of  $j_{k_l, n_{k_l}}$  and  $b_{k_l}$  remain constant. In this way we find  $j \in X$  and  $b \in \mathbb{P}(\mathbb{R}^d)$  such that

$$\frac{\mu(Z_b \cap [0; j])}{\mu([0; j])} = 1.$$

Equivalently,  $L_x = b$  for  $\mu$ -almost every  $x \in [0; j]$ . Then, iterating once more,  $L_y = A(j)b$  for  $\mu$ -almost every  $y \in X$  (the image of any full  $\mu$ -measure subset of  $[0; j]$  has full  $\mu$ -measure in  $X$ ). This implies that  $L_x$  takes a unique value on a full  $\mu$ -measure set, as claimed.  $\square$

We are ready to complete the proof of the proposition. Claim (7) follows from  $\mu(Z_{A(j)b}) = 1$  and (9). By construction,  $\tilde{\Phi}(x, v) = \lambda_+(A, P)$  for  $v \neq L$  and  $\tilde{\Phi}(x, L) < \lambda_+(A, P)$  for  $\mu$ -almost all  $x \in M$ . By the Oseledets theorem,  $\tilde{\Phi}$  only

takes the values  $\lambda_{\pm}(A, P)$ . It follows that  $\tilde{\Phi}(x, L) = \lambda_{-}(A, P)$  for  $\mu$ -almost all  $x \in M$ . This gives claim (8).  $\square$

## 5. PROBABILISTIC REPELLERS

For any  $\alpha \in \text{GL}(d)$ , we denote by  $D\alpha(v)$  the derivative at a point  $v \in \mathbb{P}(\mathbb{R}^d)$  of the action of  $\alpha$  in the projective space. The tangent space to the projective space at  $v$  is canonically identified with the hyperplane  $v^{\perp} \subset \mathbb{R}^d$  orthogonal to  $v$ . Then the derivative is given by

$$(10) \quad D\alpha(v)\dot{v} = \frac{\text{proj}_{\alpha(v)} \alpha(\dot{v})}{\|\alpha(\underline{v})\|/\|\underline{v}\|} \quad \text{for every } \dot{v} \in v^{\perp}$$

where  $\text{proj}_w : \underline{u} \mapsto \underline{u} - \frac{\underline{w}(\underline{u} \cdot \underline{w})}{(\underline{w} \cdot \underline{w})}$  denotes the orthogonal projection to the hyperplane orthogonal to  $w$ . This implies the following estimate:

$$(11) \quad \frac{1}{\|\alpha\| \|\alpha^{-1}\|} \leq \frac{\|D\alpha(v)\dot{v}\|}{\|\dot{v}\|} \leq \|\alpha\| \|\alpha^{-1}\| \quad \text{for every non-zero } \dot{v} \in v^{\perp}.$$

For  $n \geq 1$ , we will write  $DA^n(x, v)$  to mean  $D\alpha(v)$  with  $\alpha = A^n(x)$ .

We call  $v \in \mathbb{P}(\mathbb{R}^d)$  an *invariant point* for  $A$  if  $A(i)v = v$  for every  $i \in X$ . We call  $v$  a *P-expanding point* for  $A$  if it is invariant for  $A$  and there exist  $l \geq 1$  and  $c > 0$  such that

$$(12) \quad \int_{[0; i]} \log \|DA^l(x, v)^{-1}\|^{-1} d\mu(x) \geq 4cp_i \quad \text{for every } i \in X.$$

Keep in mind that  $\|\alpha^{-1}\|^{-1}$  is the *co-norm* of an (invertible) matrix  $\alpha$ :

$$\|\alpha^{-1}\|^{-1} = \inf \left\{ \frac{\|\alpha(v)\|}{\|v\|} : v \neq 0 \right\}.$$

**Remark 5.1.** If  $v$  is an invariant point for  $A$  and  $\eta$  is a  $P$ -stationary vector then

$$\eta_j(\{v\}) = \sum_{i \in X} \frac{p_i P_{i,j}}{p_j} \eta_i(\{v\}) \quad \text{for every } j \in X.$$

Since  $\sum_{i \in X} p_i P_{i,j}/p_j = 1$  for every  $j \in X$ , this implies that the value of  $\eta_j(\{v\})$  is the same for all  $j \in X$ . In particular, if  $v$  is an atom of  $\eta$  then it is an atom of  $\eta_j$ , with the same weight for every  $j \in X$ .

**Proposition 5.2.** *Let  $v \in \mathbb{P}(\mathbb{R}^d)$  be an invariant point for  $A$  and suppose that there exist  $a < b$  such that, for  $\mu$ -almost every  $x \in M$ ,*

- (1)  $\lim_n \frac{1}{n} \log \|A^n(x)\underline{v}\| \leq a$  for any  $\underline{v} \in v$
- (2)  $\lim_n \frac{1}{n} \log \|A^n(x)\underline{w}\| \geq b$  for any  $\underline{w} \notin v$ .

*Then  $v$  is a  $P$ -expanding point for  $A$ .*

*Proof.* By assumption,  $A(i)v = v$  for every  $i \in X$ . Let  $A^{\perp}$  be the linear cocycle induced by  $A$  on the orthogonal complement  $v^{\perp}$ , that is,

$$A^{\perp}(x)\dot{v} = \text{proj}_v A(x)\dot{v} \quad \text{for each } \dot{v} \in v^{\perp}.$$

The second hypothesis implies (see [17, Proposition 4.14]) that the Lyapunov exponents of  $A^{\perp}$  are bounded below by  $b$ . In other words, the smallest Lyapunov exponent

$$\lim_n \frac{1}{n} \log \|(A^{\perp})^n(x)^{-1}\|^{-1} \geq b$$

for  $\mu$ -almost every  $x \in M$ . The expression (10) implies

$$\|DA^n(x, v)^{-1}\|^{-1} \geq \frac{\|(A^\perp)^n(x)^{-1}\|^{-1}}{\|A^n(x)\underline{v}\|}$$

for any unit vector  $\underline{v} \in v$ . Then, using the first hypothesis, it follows that

$$\lim_n \frac{1}{n} \log \|DA^n(x, v)^{-1}\|^{-1} \geq b - a > 0$$

for  $\mu$ -almost every  $x \in M$ . It can be easily seen from (11) that the sequence on the left-hand side is uniformly bounded. So, we may use the bounded convergence theorem to conclude that

$$\lim_n \frac{1}{n} \int_{[0; i]} \log \|DA^n(x, v)^{-1}\|^{-1} d\mu(x) \geq (b - a)p_i$$

for every  $i \in X$ . Fix any  $c > 0$  and then take  $l \geq 1$  large enough so that

$$\int_{[0; i]} \log \|DA^l(x, v)^{-1}\|^{-1} d\mu(x) \geq l \frac{b - a}{2} p_i \geq 4cp_i$$

for every  $i \in X$ . □

**Proposition 5.3.** *Let  $v \in \mathbb{P}(\mathbb{R}^d)$  be a  $P$ -expanding point for  $A$  and  $c > 0$  and  $l \geq 1$  be as in (12). Then, assuming that  $\delta > 0$  is small enough,*

$$\int_{[0; i]} \|DA^l(x, v)^{-1}\|^\delta d\mu(x) \leq (1 - 3c\delta)p_i \quad \text{for every } i \in X.$$

*Proof.* Fix  $i \in X$ . The function

$$\delta \mapsto \psi_i(\delta) = \int_{[0; i]} \|DA^l(x, v)^{-1}\|^\delta d\mu(x)$$

is differentiable and the derivative is given by

$$(13) \quad \psi'_i(\delta) = \int_{[0; i]} \|DA^l(x, v)^{-1}\|^\delta \log \|DA^l(x, v)^{-1}\| d\mu(x).$$

In particular, by the inequality (12),

$$\psi'_i(0) = \int_{[0; i]} \log \|DA^l(x, v)^{-1}\| d\mu(x) \leq -4cp_i.$$

Now, the factor  $\|DA^l(x, v)^{-1}\|^\delta$  in (13) is close to 1, uniformly in  $x$  and  $v$ , if  $\delta$  is close to zero. It follows that  $\psi'_i(\delta) \leq -3cp_i$  for every small  $\delta > 0$ . Since  $\psi_i(0) = p_i$ , this gives the claim. □

## 6. PROOF OF THEOREM A

The core of the proof of Theorem A is the following statement, whose proof will be given in Sections 7 through 9:

**Theorem 6.1.** *Suppose that  $v \in \mathbb{P}(\mathbb{R}^d)$  is a  $P$ -expanding point for  $A$ . Let  $(A_k, P_k)_k$  be a sequence converging to  $(A, P)$  and, for each  $k \geq 1$ , let  $\eta_k$  be a  $P_k$ -stationary unit vector. Suppose that  $(\eta_k)_k$  converges to some  $\eta$  in the weak\* topology. If  $v$  is an atom for  $\eta$  then  $\eta_k$  must be atomic for all large  $k$ .*

The second part of property (7) in Proposition 4.2 means that  $L$  is an invariant point for  $A$ . In view of property (8) in that same proposition, Proposition 5.2 implies that  $L$  is also a  $P$ -expanding point for  $A$ . By Theorem 6.1 applied to this point,  $\eta_k$  has some atom for every large  $k$ . Then, by Proposition 3.4, there exists a finite set  $\mathcal{L}_k \subset \mathbb{P}(\mathbb{R}^2)$  such that  $A_k(i)v \in \mathcal{L}_k$  and  $\eta_{k,i}(\{v\}) > 0$  for every  $v \in \mathcal{L}_k$  and every  $i \in X$ .

We claim that  $\#\mathcal{L}_k \leq 2$  if  $k$  is large. That can be seen as follows. The existence of an invariant subspace  $L$  implies that no matrix  $A(i)$  is elliptic. If all the  $A(i)$  were either parabolic matrices (necessarily with the same eigenspace) or multiples of the identity, then we would have  $\lambda_-(A, P) = \lambda_+(A, P)$ , which is assumed not to be the case. Thus, there exists  $i \in X$  such that  $A(i)$  is hyperbolic. Since the set of hyperbolic matrices is open, it follows that  $A_k(i)$  is hyperbolic for all large  $k$ . Now it suffices to notice that for a hyperbolic  $2 \times 2$  matrix, any finite invariant set consists of either one or two eigenspaces.

Suppose that  $\mathcal{L}_k$  has a single element  $L_k$ . Define  $\Phi_k : M \times \mathbb{P}(\mathbb{R}^2) \rightarrow \mathbb{R}$  by  $\Phi_k(x, v) = \log(\|A_k(x)\underline{v}\|/\|\underline{v}\|)$ . We claim that

$$(14) \quad \int \Phi_k(x, L_k) d\mu(x) = \lambda_+(A_k, P_k).$$

To prove this, let  $\zeta_k = (\zeta_{k,i})_{i \in X}$ , where  $\zeta_{k,i}$  is the Dirac mass at the point  $L_k$ , for every  $i \in X$ . Observe that  $\zeta_k$  is  $P_k$ -stationary, and so, using the ergodic theorem for  $\mu_k \times \zeta_k$ ,

$$(15) \quad \int \Phi_k(x, L_k) d\mu_k(x) = \int \tilde{\Phi}_k(x, L_k) d\mu_k(x).$$

We also have that  $\tilde{\Phi}_k(x, v) = \lambda_+(A_k, P_k)$  for  $(\mu_k \times \eta_k)$ -almost every  $(x, v)$ , because

- $\tilde{\Phi}_k(x, v) \leq \lambda_+(A_k, P_k)$  for  $(\mu_k \times \eta_k)$ -almost every  $(x, v)$ ;
- our choice of  $\eta_k$  means that  $\int \tilde{\Phi}_k d(\mu_k \times \eta_k) = \lambda_+(A_k, P_k)$ .

Since  $L_k$  is an atom for  $\eta_k$ , it follows that  $\tilde{\Phi}_k(x, L_k) = \lambda_+(A_k, P_k)$  for  $\mu_k$ -almost every  $x \in M$ . Together with (15), this implies (14).

Up to restricting to a subsequence, we may assume that  $(L_k)_k$  converges to some  $L_0 \in \mathbb{P}(\mathbb{R}^2)$ . Then  $A(i)L_0 = L_0$  for every  $i \in X$ . Let  $\zeta_0 = (\zeta_{0,i})_{i \in X}$  where  $\zeta_{0,i}$  is the Dirac mass at  $L_0$  for every  $i \in X$ . Note that  $\zeta_0$  is  $P$ -stationary, and so the ergodic theorem gives that

$$(16) \quad \int \Phi(x, L_0) d\mu(x) = \int \tilde{\Phi}(x, L_0) d\mu(x).$$

Clearly,  $(\Phi_k)_k$  converges uniformly to  $k$  when  $k \rightarrow \infty$ . So, (14) yields

$$(17) \quad \lambda_+(A_k, P_k) \rightarrow \int \Phi(x, L_0) d\mu(x).$$

According to Proposition 4.2, there are two possibilities:

- If  $L_0 \neq L$  then the right-hand side of (16) is equal to  $\lambda_+(A, P)$ . Then (17) means that  $\lambda_+(A_k, P_k)$  converges to  $\lambda_+(A, P)$ , which contradicts (6).
- If  $L_0 = L$  then the right-hand side of (16) is equal to  $\lambda_-(A, P)$ . Then  $\lambda_+(A_k, P_k) \rightarrow \lambda_-(A, P)$  and, by Lemma 4.1,  $\lambda_-(A_k, P_k) \rightarrow \lambda_+(A, P)$ . This is a contradiction, as  $\lambda_- \leq \lambda_+$  and the inequality is strict at  $(A, P)$ .

We have shown that, for  $k$  large,  $\mathcal{L}_k$  can not consist of a single point.

Now suppose that  $\mathcal{L}_k$  consists of two points,  $L_k$  and  $L'_k$ . The arguments are similar. Let  $\zeta_k = (\zeta_{k,i})_{i \in X}$  with  $\zeta_{k,i} = (\delta_{L_k} + \delta_{L'_k})/2$ . Arguing as we did for (14), we find that

$$(18) \quad \int \frac{1}{2} (\Phi_k(x, L_k) + \Phi_k(x, L'_k)) d\mu_k(x) = \lambda_+(A_k, P_k).$$

We may assume that  $(L_k)_k$  and  $(L'_k)_k$  converge to subspaces  $L_0$  and  $L'_0$ , respectively. Let  $\zeta_0 = (\zeta_{0,i})_{i \in X}$  with  $\zeta_{0,i} = (\delta_{L_0} + \delta_{L'_0})/2$ . Then

$$(19) \quad \int \frac{1}{2} (\Phi(x, L_0) + \Phi(x, L'_0)) d\mu(x) = \int \frac{1}{2} (\tilde{\Phi}(x, L_0) + \tilde{\Phi}(x, L'_0)) d\mu(x)$$

and, taking the limit in (18),

$$\lambda_+(A_k, P_k) \rightarrow \int \frac{1}{2} (\Phi(x, L_0) + \Phi(x, L'_0)) d\mu(x).$$

If  $L_0$  and  $L'_0$  are both different from  $L$  then the right-hand side of (19) is equal to  $\lambda_+(A, P)$  and we reach a contradiction just as we did in case (i) of the previous paragraph. Now suppose that  $L_0 = L$ . We may assume that, for each  $k$  large, there exists  $i_k \in X$  such that  $A_k(i_k)L_k = L'_k$ ; otherwise, we could take  $\mathcal{L}_k = \{L_k\}$  instead, and that case has already been dealt with. Passing to the limit along a convenient subsequence, we find  $i_0 \in X$  such that  $A(i_0)L_0 = L'_0$ . By the second part of (7), this implies that  $L'_0$  is also equal to  $L$ . Then the right-hand side of (19) is equal to  $\lambda_-(A, P)$ , and we reach a contradiction just as we did in case (ii) of the previous paragraph. So, for  $k$  large,  $\mathcal{L}_k$  can not consist of two points either.

This reduces the proof of Theorem A to proving Theorem 6.1.

## 7. COUPLINGS AND ENERGY

In this section we prepare the proof of Theorem 6.1. Let  $v$  be a  $P$ -expanding point for  $A$  and assume that it is an atom for  $\eta$ . According to Remark 5.1, there exists  $\kappa > 0$  such that  $\eta_i(\{v\}) = \kappa$  for every  $i \in X$ . Let  $U \subset \mathbb{P}(\mathbb{R}^d)$  be an open neighborhood of  $v$  such that

$$\eta_i(\bar{U}) < \frac{10}{9}\kappa \quad \text{for every } i \in X.$$

Fix constants  $c > 0$ ,  $l \geq 1$  and  $\delta > 0$  as in Proposition 5.3 and take  $k$  to be large enough that the conclusion of the proposition holds (further conditions will be imposed on  $k$  along the way). Since  $A^l$  is continuous and  $A_k^l$  converges to  $A^l$  (uniformly) when  $k \rightarrow \infty$ , there exists an open neighborhood  $U_1 \subset U$  of  $v$  such that

$$(20) \quad \sup_{z \in U_1} \|DA_k^l(x, z)^{-1}\|^\delta \leq (1 + c\delta) \|DA^l(x, v)^{-1}\|^\delta$$

for every  $x \in M$  and every large  $k$ . Reducing  $U_1$  if necessary, we may also assume that

$$(21) \quad A_k^l(x)^{-1}(\bar{U}_1) \subset U \quad \text{for every } x \in M \text{ and every } k \text{ large.}$$

Let  $d$  be the distance on  $\mathbb{P}(\mathbb{R}^d)$  defined by the angle between two directions, normalized in such a way that the diameter is 1. By the mean value theorem and

(20),

$$\begin{aligned} d(u, w)^\delta &\leq d(A_k^l(x)u, A_k^l(x)w)^\delta \sup_{z \in U_1} \|DA_k^l(x, z)^{-1}\|^\delta \\ &\leq d(A_k^l(x)u, A_k^l(x)w)^\delta (1 + c\delta) \|DA^l(x, v)^{-1}\|^\delta \end{aligned}$$

for any pair of distinct points  $u, w \in U_1$ , any  $x \in M$  and any large  $k$ . Then,

$$\begin{aligned} \int_{[0; i]} d(A_k^l(x)u, A_k^l(x)w)^{-\delta} d\mu_k(x) \\ \leq (1 + c\delta) d(u, w)^{-\delta} \int_{[0; i]} \|DA^l(x, v)^{-1}\|^\delta d\mu_k(x) \\ \leq (1 + 2c\delta) d(u, w)^{-\delta} \int_{[0; i]} \|DA^l(x, v)^{-1}\|^\delta d\mu(x) \end{aligned}$$

(the last inequality uses the fact that  $\mu_k \rightarrow \mu$  and assumes that  $k$  is large enough). Using Proposition 5.3, we conclude that

$$\begin{aligned} (22) \quad \int_{[0; i]} d(A_k^l(x)u, A_k^l(x)w)^{-\delta} d\mu_k(x) &\leq (1 + 2c\delta) d(u, w)^{-\delta} (1 - 3c\delta) p_{k, i} \\ &\leq (1 - c\delta) p_{k, i} d(u, w)^{-\delta} \end{aligned}$$

for any  $i \in X$ , any pair of distinct points  $u, w \in U_1$ , and any large  $k$ .

For any measure  $\xi$  on  $\mathbb{P}(\mathbb{R}^d)^2$  and any  $\delta > 0$ , define the  $\delta$ -energy of  $\xi$  to be

$$E_\delta(\xi) = \int \Psi d\xi,$$

where

$$\Psi(u, w) = \begin{cases} d(u, w)^{-\delta} & \text{if } (u, w) \in U_1 \times U_1 \\ 1 & \text{otherwise.} \end{cases}$$

Let  $\pi_s : \mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{P}(\mathbb{R}^d)$  be the projection on the  $s$ -th coordinate, for  $s = 1, 2$ . The *mass*  $\|\eta\|$  of a measure  $\eta$  on  $\mathbb{P}(\mathbb{R}^d)$  is defined by  $\|\eta\| = \eta(\mathbb{P}(\mathbb{R}^d))$ . If  $\eta_1$  and  $\eta_2$  are measures on  $\mathbb{P}(\mathbb{R}^d)$  with the same mass, a *coupling* of  $\eta_1$  and  $\eta_2$  is a measure  $\xi$  on  $\mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d)$  such that  $(\pi_s)_*\xi = \eta_s$  for  $s = 1, 2$ . Define:

$$e_\delta(\eta_1, \eta_2) = \inf\{E_\delta(\xi) : \xi \text{ a coupling of } \eta_1 \text{ and } \eta_2\}.$$

A *self-coupling* of a measure  $\eta$  is a coupling of  $\eta_1 = \eta$  and  $\eta_2 = \eta$ . We call a self-coupling *symmetric* if it is invariant under the involution  $\iota : (x, y) \mapsto (y, x)$ . Define the  $\delta$ -energy of  $\eta$  to be

$$(23) \quad e_\delta(\eta) = e_\delta(\eta, \eta) = \inf\{E_\delta(\xi) : \xi \text{ a self-coupling of } \eta\}.$$

**Remark 7.1.** The function  $\xi \mapsto E_\delta(\xi)$  is lower semi-continuous. Indeed, it is not difficult to find bounded continuous functions  $\Psi_n : \mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$  increasing to  $\Psi$  as  $n \rightarrow \infty$ . Then, given  $(\xi_k)_k \rightarrow \xi$  and  $\varepsilon > 0$ , we may fix  $n$  such that  $\int \Psi_n d\xi \geq \int E_\delta(\xi) - \varepsilon$  (if the energy is infinite, replace the right-hand side by  $\varepsilon^{-1}$ ). Then, by the definition of weak\* topology,

$$\liminf_k E_\delta(\xi_k) \geq \liminf_k \int \Psi_n d\xi_k \geq \int \Psi_n d\xi \geq E_\delta(\xi) - \varepsilon$$

( $\geq \varepsilon^{-1}$  if the energy is infinite). Making  $\varepsilon \rightarrow 0$  one gets lower semi-continuity.

As a consequence, the infimum in (23) is always achieved. Moreover, that remains true if we restrict to symmetric self-couplings: just note that if  $\xi$  is a self-coupling of  $\eta$  then  $\xi' = (\xi + \iota_*\xi)/2$  is a symmetric self-coupling of  $\eta$  and  $E_\delta(\xi) = E_\delta(\xi')$ . As

a further consequence, the function  $\eta \mapsto e_\delta(\eta)$  is lower semi-continuous. Indeed, given  $(\eta_k)_k \rightarrow \eta$ , take self-couplings  $(\xi_k)_k$  with  $E_\delta(\xi_k) = e_\delta(\eta_k)$ . Up to restricting to a subsequence,  $(\xi_k)_k$  converges to some  $\xi$ , which is necessarily a self-coupling of  $\eta$ , and

$$e_\delta(\eta) \leq E_\delta(\xi) \leq \liminf_k E_\delta(\xi_k) = \liminf_k e_\delta(\eta_k).$$

This proves lower semi-continuity.

**Lemma 7.2.** *For any measure  $\eta$  in  $\mathbb{P}(\mathbb{R}^d)$  and any  $\delta > 0$ :*

- (1) *If  $\eta(\{u\}) > \|\eta\|/2$  for some  $u \in U_1$  then  $e_\delta(\eta) = \infty$ .*
- (2) *If  $\eta(\{u\}) < \|\eta\|/2$  for every  $u \in \bar{U}_1$  then  $e_\delta(\eta) < \infty$ .*

*Proof.* First we prove (1). Take  $u$  as in the hypothesis. For any self-coupling  $\xi$  of  $\eta$ , we have

$$\xi(\{u\}^c \times \mathbb{P}(\mathbb{R}^d)) = \eta(\{u\}^c) = \xi(\mathbb{P}(\mathbb{R}^d) \times \{u\}^c).$$

Taking the union,  $\xi(\{(u, u)\}^c) \leq 2\eta(\{u\}^c) < \eta(\mathbb{P}(\mathbb{R}^d)) = \xi(\mathbb{P}(\mathbb{R}^d)^2)$ . This implies that  $(u, u)$  is an atom for  $\xi$ . Hence  $E_\delta(\xi) = \infty$ , as claimed in (1).

A result analogous to (2) was proven in [17, Lemma 10.7], for a slightly different notion of energy. Exactly the same arguments can be used in the present situation to construct a self-coupling  $\xi$  of  $\eta$  which gives zero weight to a neighborhood of  $\{(u, u) : u \in \bar{U}_1\}$  and, hence, has finite  $\delta$ -energy.  $\square$

Let  $\eta = (\eta_i)_{i \in X}$  be a measure vector on  $\mathbb{P}(\mathbb{R}^d)$  and  $\xi = (\xi_i)_{i \in X}$  be a measure vector on  $\mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d)$ . We say that  $\xi$  is a (symmetric) self-coupling of  $\eta$  if  $\xi_i$  is a (symmetric) self-coupling of  $\eta_i$  for every  $i \in X$ . Then we define

$$E_\delta(\xi) = \sum_{i \in X} p_i E_\delta(\xi_i)$$

and

$$e_\delta(\eta) = \inf \{E_\delta(\xi) : \xi \text{ a self-coupling of } \eta\} = \sum_{i \in X} p_i e_\delta(\eta_i)$$

(of course, this depends on the probability vector  $p$ ). It follows from Remark 7.1 that the function  $\eta \mapsto e_\delta(\eta)$  is lower semi-continuous in the space of  $P$ -stationary measure vectors. Moreover, for  $(\eta_k)_k$  as in Theorem 6.1,

$$(24) \quad \liminf_k e_\delta(\eta_k) = \liminf_k \sum_{i \in X} p_{k,i} e_\delta(\eta_{k,i}) \geq \sum_{i \in X} p_i e_\delta(\eta_i) = e_\delta(\eta).$$

## 8. THE LARGE ATOM CASE

We split the proof of Theorem 6.1 into two cases, depending on the value of  $\kappa = \eta_i(\{v\})$ . In this section we treat the case  $\kappa > 9/10$ .

**Proposition 8.1.** *There exists  $C > 0$  such that, for every large  $k$ , given any symmetric self-coupling  $\xi_k$  of  $\eta_k$ , there exists a symmetric self-coupling  $\xi_k''$  of  $\eta_k$  satisfying*

$$E_\delta(\xi_k'') \leq C + (1 - c\delta)E_\delta(\xi_k).$$

The present case of Theorem 6.1 can be deduced as follows. Suppose that there exists a subsequence  $(k_j)_j \rightarrow \infty$  such that  $\eta_{k_j}$  has no atoms and, hence,  $e_\delta(\eta_{k_j})$  is finite. Then, Proposition 8.1 gives that

$$e_\delta(\eta_{k_j}) \leq C/(c\delta) \quad \text{for every large } j.$$



By the lower semi-continuity property (24), it follows that  $e_\delta(\eta) \leq C/(c\delta)$ . However, by Lemma 7.2, the assumption  $\kappa > 9/10$  implies that  $e_\delta(\eta) = \infty$ . This contradiction proves that  $\eta_k$  does have some atom for every large  $k$ .

In the remainder of this section we prove Proposition 8.1. Let  $\xi_k$  be a self-coupling of  $\eta_k$ . The construction of the self-coupling  $\xi_k''$  will be done in two steps. Fix open neighborhoods  $U_3 \subset \bar{U}_3 \subset U_2 \subset \bar{U}_2 \subset U_1$  of the point  $v$ . The first step is:

**Lemma 8.2.** *There exists  $C_1 > 0$  such that, assuming  $k$  is sufficiently large, given any self-coupling  $\xi_k$  of  $\eta_k$  there exists a self-coupling  $\xi_k'$  of  $\eta_k$  satisfying*

- (a)  $\xi_{k,i}'(U_2^c \times U_2^c) = 0$  for every  $i \in X$ ;
- (b)  $E_\delta(\xi_{k,i}') \leq E_\delta(\xi_{k,i}) + C_1$  for every  $i \in X$ .

*Proof.* For each  $i \in X$ , let  $\zeta_{k,i}$  be the projection of  $(\xi_{k,i} \mid U_2^c \times U_2^c)$  and  $\eta_{k,i}^3$  be the projection of  $(\xi_{k,i} \mid U_3 \times U_3)$ , on either coordinate. Then take

$$\begin{aligned} \xi_{k,i}' &= \xi_{k,i} - (\xi_{k,i} \mid U_2^c \times U_2^c) - \frac{\|\zeta_{k,i}\|}{\|\eta_{k,i}^3\|} (\xi_{k,i} \mid U_3 \times U_3) \\ &\quad + \frac{1}{\|\eta_{k,i}^3\|} [\zeta_{k,i} \times \eta_{k,i}^3 + \eta_{k,i}^3 \times \zeta_{k,i}] \end{aligned}$$

Assuming  $k$  is sufficiently large,  $\eta_{k,i}(U_3) > (8/9)\kappa > 8/10$  for all  $i \in X$ . Then,  $\|\xi_{k,i} \mid U_3 \times U_3\| > 6/10$ . It follows that

$$\|\eta_{k,i}^3\| = \|\xi_{k,i} \mid U_3 \times U_3\| > \frac{6}{10} \quad \text{and} \quad \|\zeta_{k,i}\| \leq \|\eta_{k,i} \mid U_2^c\| < \frac{2}{10}.$$

In particular,  $\|\zeta_{k,i}\| < \|\eta_{k,i}^3\|$ , which ensures that  $\xi_{k,i}'$  is a positive measure. Next, notice that the projection of  $\xi_{k,i}'$  on either coordinate is equal to

$$\eta_{k,i} - \zeta_{k,i} - \frac{\|\zeta_{k,i}\|}{\|\eta_{k,i}^3\|} \eta_{k,i}^3 + \zeta_{k,i} + \frac{\|\zeta_{k,i}\|}{\|\eta_{k,i}^3\|} \eta_{k,i}^3 = \eta_{k,i}.$$

So,  $\xi_{k,i}'$  is indeed a self-coupling of  $\eta_{k,i}$ . Finally, it is clear that

$$(25) \quad E_\delta(\xi_{k,i}') \leq E_\delta(\xi_{k,i}) + \frac{1}{\|\eta_{k,i}^3\|} E_\delta(\zeta_{k,i} \times \eta_{k,i}^3 + \eta_{k,i}^3 \times \zeta_{k,i}).$$

The measure  $(\zeta_{k,i} \times \eta_{k,i}^3 + \eta_{k,i}^3 \times \zeta_{k,i})$  is concentrated on the set  $(U_2^c \times U_3 \cup U_3 \times U_2^c)$ , which is at positive distance from the diagonal. Thus, its energy is uniformly bounded. We have already seen that  $\|\eta_{k,i}^3\|$  is bounded away from zero. Thus, the last term in (25) is uniformly bounded, and that yields the claim in the lemma.  $\square$

Let  $\mathcal{P}_k$  denote the operator defined as in (2), with  $(A, P)$  replaced with  $(A_k, P_k)$ . The second, and final step of our construction uses the *diagonal action*  $\mathfrak{P}_k$  of  $\mathcal{P}_k$ , defined as follows:

$$(\mathfrak{P}_k \xi)_j(D_1 \times D_2) = \sum_{i \in X} \frac{p_{k,i} P_{k,i,j}}{p_{k,j}} \xi_i(A_k(i)^{-1}(D_1) \times A_k(i)^{-1}(D_2))$$

for any  $j \in X$  and any measurable  $D \subset \mathbb{P}(\mathbb{R}^d)$ . Equivalently,

$$(26) \quad p_{k,j} \int \Phi d(\mathfrak{P}_k \xi)_j = \sum_{i \in X} \int_{[0;i,j]} \int \Phi(A_k(x)u, A_k(x)w) d\xi_i(u, w) d\mu_k(x)$$

for any measurable function  $\Phi : \mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d) \rightarrow [0, \infty)$  (the two sides of the equality may be infinite). It is clear that

$$(27) \quad (\pi_s)_* \circ \mathfrak{P}_k = \mathcal{P}_k \circ (\pi_s)_* \quad \text{for } s = 1, 2.$$

Take  $\xi_k'' = \mathfrak{P}_k^l \xi_k'$ . Notice that  $\xi_k''$  is a symmetric measure and a self-coupling of  $\eta_k$ : the previous relation gives that

$$(\pi_s)_* \xi_k'' = \mathcal{P}_k^l((\pi_s)_* \xi_k') = \mathcal{P}_k^l \eta_k = \eta_k \quad \text{for } s = 1, 2.$$

**Lemma 8.3.** *There exists  $C_2 > 0$  such that, assuming  $k$  is sufficiently large,*

$$E_\delta(\xi_k'') \leq (1 - c\delta)E_\delta(\xi_k') + C_2.$$

*Proof.* From the definition of  $\mathfrak{P}_k$  in (26) we get that

$$p_{k,j} \int \Psi d\xi_{k,j}'' = \sum_{i \in X} \int_{[0;i] \cap [l;j]} \int \Psi(A_k^l(x)u, A_k^l(x)w) d\xi_{k,i}'(u, w) d\mu_k(x).$$

for every  $j \in X$ . Thus, adding over  $j$ ,

$$(28) \quad E_\delta(\xi_k'') = \sum_{i \in X} \int_{[0;i]} \int \Psi(A_k^l(x)u, A_k^l(x)w) d\xi_{k,i}'(u, w) d\mu_k(x).$$

If  $(u, w) \in (U_1 \times U_1)$ , then  $\Psi(u, w) = d(u, w)^{-\delta}$ . Using (22), we get that

$$(29) \quad \begin{aligned} \int_{[0;i]} \Psi(A_k^l(x)u, A_k^l(x)w) d\mu_k(x) &\leq \int_{[0;i]} d(A_k^l(x)u, A_k^l(x)w)^{-\delta} d\mu_k(x) \\ &\leq (1 - c\delta) p_{k,i} \Psi(u, w). \end{aligned}$$

Now suppose that  $(u, w) \notin (U_1 \times U_1)$ . By the property in Lemma 8.2(a), we only need to consider  $(u, w) \notin (U_2^c \times U_2^c)$ . Then  $(u, w) \in U_2 \times U_1^c \cup U_1^c \times U_2$ . Since the latter set is at positive distance from the diagonal, it follows that  $d(u, w)$  is uniformly bounded from below. Hence, there exists  $C_2 > 0$  such that

$$(30) \quad \Psi(A_k^l(x)u, A_k^l(x)w) \leq d(A_k^l(x)u, A_k^l(x)w)^{-\delta} \leq C_2.$$

for every  $x \in M$ . Using (29) and (30), together with Fubini, we get that (28) yields

$$E_\delta(\xi_k'') \leq \sum_{i \in X} \int_{I_1} (1 - c\delta) p_{k,i} \Psi(u, w) d\xi_{k,i}'(u, w) + \int_{I_1^c} C_2 p_{k,i} d\xi_{k,i}'(u, w)$$

where  $I_1 = \{(u, w) \in U_1 \times U_1\}$ . Then,

$$E_\delta(\xi_k'') \leq (1 - c\delta) \sum_{i \in X} p_{k,i} E_\delta(\xi_{k,i}') + C_2 p_{k,i} = (1 - c\delta) E_\delta(\xi_k') + C_2$$

as claimed.  $\square$

Combining Lemmas 8.2 and 8.3, we find that

$$E_\delta(\xi_k'') \leq (1 - c\delta)E_\delta(\xi_k) + C \quad \text{with } C = (1 - c\delta)C_1 + C_2.$$

This finishes the proof of Proposition 8.1.

## 9. THE SMALL ATOM CASE

We are left to consider  $\kappa \leq 9/10$ . This is similar to the previous case, but the analysis must be localized on suitable neighborhoods of  $v$ , inside which the relative weight of the atom is close to 1, and the precise choice of such neighborhoods turns out to be much more delicate than one would anticipate. We do as follows.

Since  $\kappa \leq 9/10$ , we have  $\eta_{k,i}(\bar{U}) < (10/9)\kappa \leq 1$  for every  $i \in X$  and every large  $k$ . Suppose that  $\eta_k$  is non-atomic for arbitrarily large values of  $k$ . Up to restricting to a subsequence, we may suppose that this is the case for every  $k$ . Then every  $\eta_{k,i}$  is a continuous measure and so we may find  $V_{k,i} \supset \bar{U}$  such that  $\eta_{k,i}(V_{k,i}) = (10/9)\kappa$ . We denote

$$V_k = (V_{k,i})_{i \in X} \quad \text{and} \quad \eta_k \mid V_k = (\eta_{k,i} \mid V_{k,i})_{i \in X} \quad \text{for each } k \geq 1.$$

The construction we have just described is designed to get the following fact, which will be needed in a while:

**Lemma 9.1.**  $\mathcal{P}_k^l(\eta_k \mid V_k^c)_j(V_{k,j}) = \mathcal{P}_k^l(\eta_k \mid V_k)_j(V_{k,j}^c)$  for every  $j \in X$ .

*Proof.* Clearly,

$$(31) \quad \begin{aligned} \mathcal{P}_k^l(\eta_k \mid V_k)_j(\mathbb{P}(\mathbb{R}^2)) &= \mathcal{P}_k^l(\eta_k \mid V_k)_j(V_{k,j}) + \mathcal{P}_k^l(\eta_k \mid V_k)_j(V_{k,j}^c) \\ \eta_{k,j}(V_{k,j}) &= (\mathcal{P}_k^l \eta_k)_j(V_{k,j}) = \mathcal{P}_k^l(\eta_k \mid V_k)_j(V_{k,j}) + \mathcal{P}_k^l(\eta_k \mid V_k^c)_j(V_{k,j}). \end{aligned}$$

Now, the definition of  $\mathcal{P}_k$  in (2) gives that

$$\begin{aligned} \mathcal{P}_k^l(\eta_k \mid V_k)_j(\mathbb{P}(\mathbb{R}^2)) &= \sum_{i \in X} \frac{1}{p_{k,j}} \int_{[0;i] \cap [l;j]} \eta_{k,i}(A_k^l(x)^{-1}(\mathbb{P}(\mathbb{R}^2)) \cap V_{k,i}) d\mu_k(x) \\ &= \sum_{i \in X} \frac{1}{p_{k,j}} \int_{[0;i] \cap [l;j]} \eta_{k,i}(V_{k,i}) d\mu_k(x). \end{aligned}$$

The key observation is that, by construction,  $\eta_{k,i}(V_{k,i})$  is independent of  $i$  and  $k$ . Thus the previous equality may be rewritten as

$$(32) \quad \begin{aligned} \mathcal{P}_k^l(\eta_k \mid V_k)_j(\mathbb{P}(\mathbb{R}^2)) &= \sum_{i \in X} \frac{1}{p_{k,j}} \int_{[0;i] \cap [l;j]} \eta_{k,j}(V_{k,j}) d\mu_k(x) \\ &= \frac{1}{p_{k,j}} \int_{[l;j]} \eta_{k,j}(V_{k,j}) d\mu_k(x) = \eta_{k,j}(V_{k,j}). \end{aligned}$$

The claim is a direct consequence of (31) and (32).  $\square$

We are going to prove the following localized version of Proposition 8.1:

**Proposition 9.2.** *There exists  $C > 0$  such that, for every large  $k$ , given any symmetric self-coupling  $\xi_k$  of  $\eta_k \mid V_k$ , there exists a symmetric self-coupling  $\xi_k''$  of  $\eta_k \mid V_k$  satisfying*

$$E_\delta(\xi_k'') \leq C + (1 - c\delta)E_\delta(\xi_k).$$

The present case of Theorem 6.1 may be deduced as follows. Since the  $\eta_k$  have no atoms, the energy  $e_\delta(\eta_k \mid V_k)$  is finite. Then, Proposition 9.2 gives that

$$e_\delta(\eta_k \mid V_k) \leq C/(c\delta) \quad \text{for every large } j.$$

Let  $\hat{\eta} = (\hat{\eta}_i)_{i \in X}$  be any accumulation point of  $(\eta_k | V_k)$ . Then  $e_\delta(\hat{\eta}) \leq C/(c\delta)$ , by the lower semi-continuity property (24). For any small open neighborhood  $W$  of  $L$  and any  $i \in X$ ,

$$\hat{\eta}_i(\bar{W}) \geq \limsup_k \eta_{k,i}(V_{k,i} \cap \bar{W}) = \limsup_k \eta_{k,i}(\bar{W}) \geq \liminf_k \eta_{k,i}(W) \geq \eta_i(W)$$

because  $\eta_{k,i} | V_{k,i}$  converges to  $\hat{\eta}_i$  and  $\eta_{k,i}$  converges to  $\eta_i$  in the weak\* topology. Making  $W \rightarrow \{L\}$ , we obtain that  $\hat{\eta}_i(\{L\}) \geq \kappa = (9/10)\|\hat{\eta}_i\|$  for all  $i \in X$ . This implies that  $e_\delta(\hat{\eta}) = \infty$ , contradicting the previous conclusion. This contradiction proves that  $\eta_k$  does have some atom for all  $k$  sufficiently large.

In the remainder of this section we prove Proposition 9.2. Let  $U, U_1, U_2, U_3$  be as before. The first step is:

**Lemma 9.3.** *There exists  $C_1 > 0$  such that, assuming  $k$  is sufficiently large, given any self-coupling  $\xi_k$  of  $\eta_k | V_k$  there exists a self-coupling  $\xi'_k$  of  $\eta_k | V_k$  satisfying*

- (a)  $\xi'_{k,i}((V_{k,i} \setminus U_2) \times (V_{k,i} \setminus U_2)) = 0$  for every  $i \in X$ ;
- (b)  $E_\delta(\xi'_{k,i}) \leq E_\delta(\xi_{k,i}) + C_1$  for every  $i \in X$ .

*Proof.* Analogous to Lemma 8.2, with  $V_{k,i} \setminus U_2$  in the place of  $U_2^c$ .  $\square$

We also have the following variation of Lemma 8.3:

**Lemma 9.4.** *There exists  $C_2 > 0$  such that, assuming  $k$  is sufficiently large,*

$$E_\delta(\mathfrak{P}_k^l \xi'_k) \leq (1 - c\delta)E_\delta(\xi'_k) + C_2.$$

*Proof.* Analogous to Lemma 8.3, with  $V_{k,i} \setminus U_2$  in the place of  $U_2^c$ .  $\square$

Now we come to a main difference with respect to the more global situation treated in Section 8. Using (27) one gets that

$$\tilde{\xi}_k = \mathfrak{P}_k^l \xi'_k \text{ is a symmetric self-coupling of } \mathcal{P}_k^l(\eta_k | V_k).$$

However, this *does not* mean that  $\tilde{\xi}_k$  is a self-coupling of  $(\eta_k | V_k)$  as the latter measure need not be  $P_k$ -stationary.

To bypass this difficulty, we begin by relating  $(\eta_k | V_k)$  to the projection of  $\tilde{\xi}_k$  or, more precisely, to

$$\eta_k^1 = \text{projection of } (\tilde{\xi}_k | V_k \times V_k).$$

The next lemma shows that  $\eta_k^1 \leq (\eta_k | V_k)$  and describes the difference between the two vectors.

**Lemma 9.5.**  $(\eta_k | V_k) = \eta_k^1 + I_k + O_k$  with

- (1)  $I_k = \text{restriction of } \mathcal{P}_k^l(\eta_k | V_k^c) \text{ to } V_k$ ;
- (2)  $O_k = \text{is the projection of } (\tilde{\xi}_k | V_k \times V_k^c) \text{ on the first coordinate}$ ;
- (3)  $\|O_{k,j}\| \leq \|I_{k,j}\| \leq (2/9)\kappa$  for every  $j \in X$  and  $k$  sufficiently large.

*Proof.* On the one hand,

$$\pi_{1*}\tilde{\xi}_k = \mathcal{P}_k^l(\eta_k | V_k) = \mathcal{P}_k^l(\eta_k) - \mathcal{P}_k^l(\eta_k | V_k^c) = \eta_k - \mathcal{P}_k^l(\eta_k | V_k^c)$$

because  $\eta_k$  is  $\mathcal{P}_k^l$ -stationary. Restricting to  $V_k$  we get that

$$(\pi_{1*}\tilde{\xi}_k) | V_k = (\eta_k | V_k) - I_k.$$

On the other hand,

$$(\pi_{1*}\tilde{\xi}_k) | V_k = \pi_{1*}(\tilde{\xi}_k | V_k \times V_k) + \pi_{1*}(\tilde{\xi}_k | V_k \times V_k^c) = \eta_k^1 + O_k.$$

Combining two equalities one gets that  $(\eta_k | V_k) = \eta_k^1 + I_k + O_k$ .

We are left to proving the estimates in part (3) of the statement. Let  $j \in X$  be fixed. By Lemma 9.1,

$$\|I_{k,j}\| = \mathcal{P}_k^l(\eta_k | V_k^c)_j(V_{k,j}) = \mathcal{P}_k^l(\eta_k | V_k)_j(V_{k,j}^c).$$

Since  $\tilde{\xi}_k$  is a self-coupling of  $\mathcal{P}_k^l(\eta_k | V_k)$ ,

$$\mathcal{P}_k^l(\eta_k | V_k)_j(V_{k,j}^c) = \tilde{\xi}_k(\mathbb{P}(\mathbb{R}^2) \times V_{k,j}^c) \geq \tilde{\xi}_k(V_{k,j} \times V_{k,j}^c) = \|O_{k,j}\|.$$

These two inequalities ensure that  $\|O_{k,j}\| \leq \|I_{k,j}\|$ .

Finally, the first part of the lemma implies that  $I_{k,j} \leq \eta_{k,j} | V_{k,j}$  for every  $j \in X$ . Since  $\eta_{k,j}(V_{k,j}) = (10/9)\kappa$  and  $\liminf_k \eta_{k,j}(U_1) \geq \eta_j(U_1) \geq \kappa$ , this implies that

$$I_{k,j}(V_{k,j} \setminus U_1) \leq \eta_{k,j}(V_{k,j} \setminus U_1) < \frac{2}{9}\kappa \quad \text{for every large } k.$$

The condition (21) ensures that  $A_k^l(x)^{-1}(U_1) \subset U \subset V_{k,i}$  for every  $x \in [0; i]$ , every  $i \in X$  and every large  $k$ . Thus,

$$\begin{aligned} I_{k,j}(U_1) &= \mathcal{P}_k^l(\eta_k | V_k^c)_j(U_1) \\ &= \sum_{i \in X} \frac{1}{p_{k,j}} \int_{[0;i] \cap [l;j]} \eta_{k,i}(V_{k,i}^c \cap A_k^l(x)^{-1}(U_1)) d\mu_k(x) = 0 \end{aligned}$$

for every large  $k$ . These two facts imply that  $\|I_{k,j}\| = I_{k,j}(V_{k,j} \setminus U_1) \leq (2/9)\kappa$ , as claimed.  $\square$

According to Lemma 9.5, the difference  $(\eta_k | V_k) - \eta_k^1$  is positive and relatively small. This suggests that we try and construct the self-coupling  $\xi_k''$  of  $(\eta_k | V_k)$  we are looking for by adding suitable correcting terms to the restriction of  $\tilde{\xi}_k$  to the  $V_k \times V_k$ . These correcting terms should be concentrated outside a neighborhood of the diagonal, if possible, so that their contribution to the total energy is bounded. We choose  $\xi_k'' = (\xi_{k,i}'')_{i \in X}$ , with

$$\begin{aligned} \xi_{k,i}'' &= [(\tilde{\xi}_{k,i} | V_{k,i} \times V_{k,i}) - \frac{\|\zeta_{k,i}\|}{\|\eta_{k,i}^3\|}(\tilde{\xi}_{k,i} | U_3 \times U_3)] \\ (33) \quad &+ \frac{1}{\|I_{k,i}\|}[(O_{k,i} | U_2) \times I_{k,i} + I_{k,i} \times (O_{k,i} | U_2)] \\ &+ \frac{1}{\|\eta_{k,i}^3\|}[(\zeta_{k,i} \times \eta_{k,i}^3) + (\eta_{k,i}^3 \times \zeta_{k,i})] \end{aligned}$$

where  $\eta_{k,i}^3$  is the projection of  $(\tilde{\xi}_{k,i} | U_3 \times U_3)$  on either coordinate and

$$\zeta_{k,i} = (1 - \frac{\|O_{k,i} | U_2\|}{\|I_{k,i}\|})I_{k,i} + (O_{k,i} | U_2^c).$$

**Lemma 9.6.**  $\xi_k''$  is a symmetric self-coupling of  $(\eta_k | V_k)$ .

*Proof.* Let  $i \in X$  be fixed. It is clear that the three terms on the right-hand side of (33) are symmetric. Moreover,

$$\|O_{k,i} | U_2\| \leq \|O_{k,i}\| \leq \|I_{k,i}\|.$$

This ensures that  $\zeta_{k,i}$  is a positive measure. Clearly, the last two terms in (33) are positive. To see that the first term is also positive, it suffices to check that  $\|\zeta_{k,i}\| < \|\eta_{k,i}^3\|$ . That can be done as follows.

Fix an open neighborhood  $U_4 \subset U_3$  such that  $A^l(x)\bar{U}_4 \subset U_3$  for every  $x \in M$ . Assuming  $k$  is large enough,

$$\eta_{k,i}(U_4) > \frac{8}{9}\eta_i(U_4) \geq \frac{8}{9}\kappa \quad \text{for every } i \in X.$$

Since  $\eta_{k,i}(V_{k,i}) = (10/9)\kappa$ , this implies that  $\xi'_{k,i}(U_4 \times U_4) \geq (5/9)\kappa$ . Increasing  $k$  again, if necessary,  $A_k^l(x)^{-1}(U_3) \supset U_4$  for every  $x \in M$ . Hence,

$$\|\eta_{k,j}^3\| = \tilde{\xi}_{k,j}(U_3 \times U_3) \geq \sum_{i \in X} \frac{1}{p_{k,j}} \int_{[0;i] \cap [l;j]} \xi'_{k,i}(U_4 \times U_4) d\mu_k(x) \geq \frac{5}{9}\kappa.$$

So, the claim that  $\|\zeta_{k,i}\| \leq \|\eta_{k,i}^3\|$  is an immediate consequence of

$$\|\zeta_{k,i}\| \leq \|I_{k,i}\| + \|O_{k,i}\| \leq \frac{4}{9}\kappa.$$

Next, observe that the projection of  $\xi''_{k,i}$  on either coordinate is equal to

$$\left[\eta_{k,i}^1 - \frac{\|\zeta_{k,i}\|}{\|\eta_{k,i}^3\|} \eta_{k,i}^3\right] + \left[(O_{k,i} \mid U_2) + \frac{\|O_{k,i} \mid U_2\|}{\|I_{k,i}\|} I_{k,i}\right] + \left[\zeta_{k,i} + \frac{\|\zeta_{k,i}\|}{\|\eta_{k,i}^3\|} \eta_{k,i}^3\right]$$

and that adds up to  $\eta_{k,i}^1 + I_{k,i} + O_{k,i} = \eta_{k,i} \mid V_{k,i}$ . So,  $\xi''_{k,i}$  is a self-coupling of  $\eta_{k,i} \mid V_{k,i}$  as we wanted to prove.  $\square$

**Lemma 9.7.** *There exists  $C_3 > 0$  such that*

$$E_\delta(\xi''_{k,i}) \leq E_\delta(\tilde{\xi}_{k,i}) + C_3 \quad \text{for every } i \in X \text{ and } k \text{ sufficiently large.}$$

*Proof.* Notice that the supports of the two correcting terms

$$\frac{1}{\|I_{k,i}\|} [(O_{k,i} \mid U_2) \times I_{k,i} + I_{k,i} \times (O_{k,i} \mid U_2)]$$

and

$$\frac{1}{\|\eta_{k,i}^3\|} [(\zeta_{k,i} \times \eta_{k,i}^3) + (\eta_{k,i}^3 \times \zeta_{k,i})]$$

are uniformly away from the diagonal of  $\mathbb{P}(\mathbb{R}^d)^2$ . Indeed,  $O_{k,i} \mid U_2$  is concentrated on  $U_2$  while  $I_{k,i}$  is concentrated on  $U_1^c$ , as we have seen. Similarly,  $\zeta_{k,i}$  is concentrated on  $U_2^c$  while  $\eta_{k,i}^3$  is concentrated on  $U_3$ . Moreover, the masses of these two correcting terms are uniformly bounded (by 1, say). Thus, the  $\delta$ -energy of their sum is bounded by some constant  $C_3 > 0$ , for every large  $k$ . Finally, it is clear that the  $\delta$ -energy of the first term in (33) is bounded by  $E_\delta(\tilde{\xi}_{k,i} \mid V_{k,i} \times V_{k,i})$ . Therefore,

$$E_\delta(\xi''_{k,i}) \leq E_\delta(\tilde{\xi}_{k,i} \mid V_{k,i} \times V_{k,i}) + C_3 \leq E_\delta(\tilde{\xi}_{k,i}) + C_3,$$

as claimed.  $\square$

By Lemmas 9.3, 9.4 and 9.7,

$$E_\delta(\xi''_k) \leq (1 - c\delta)E_\delta(\xi_k) + (1 - c\delta)C_1 + C_2 + C_3.$$

for every large  $k$ . This proves Proposition 9.2 for  $C = (1 - c\delta)C_1 + C_2 + C_3$ .

The proof of Theorem 6.1 is complete.

## REFERENCES

- [1] J. Bochi. Genericity of zero Lyapunov exponents. *Ergod. Th. & Dynam. Sys.*, 22:1667-1696, 2002.
- [2] J. Bochi. C1-generic symplectic diffeomorphisms: partial hyperbolicity and zero centre Lyapunov exponents. *J. Inst. Math. Jussieu*, 8:49-93, 2009.
- [3] J. Bochi and M. Viana. Lyapunov exponents: how frequently are dynamical systems hyperbolic? In *Modern dynamical systems and applications*, pages 271-297. Cambridge Univ. Press, 2004.
- [4] C. Bocker. *Continuity of Lyapunov exponents for 2D random matrices*. PhD thesis, IMPA, 2009.
- [5] C. Bonatti, L. J. Díaz, and M. Viana. *Dynamics beyond uniform hyperbolicity*, volume 102 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, 2005.
- [6] J. Bourgain. Positivity and continuity of the Lyapounov exponent for shifts on  $T^d$  with arbitrary frequency vector and real analytic potential. *J. Anal. Math.*, 96:313-355, 2005.
- [7] J. Bourgain and S. Jitomirskaya. Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential. *J. Statist. Phys.*, 108:1203-1218, 2002.
- [8] H. Furstenberg. Non-commuting random products. *Trans. Amer. Math. Soc.*, 108:377-428, 1963.
- [9] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457-469, 1960.
- [10] H. Furstenberg and Yu. Kifer. Random matrix products and measures in projective spaces. *Israel J. Math.*, 10:12-32, 1983.
- [11] F. Ledrappier. Quelques propriétés des exposants caractéristiques. *Lect. Notes in Math.*, 1097:305-396, 1984.
- [12] A. M. Lyapunov. *The general problem of the stability of motion*. Taylor & Francis Ltd., 1992. Translated from Edouard Davaux's French translation (1907) of the 1892 Russian original and edited by A. T. Fuller, With an introduction and preface by Fuller, a biography of Lyapunov by V. I. Smirnov, and a bibliography of Lyapunov's works compiled by J. F. Barrett, Lyapunov centenary issue, Reprint of Internat. J. Control **55** (1992), no. 3 [MR1154209 (93e:01035)], With a foreword by Ian Stewart.
- [13] R. Mañé. Oseledec's theorem from the generic viewpoint. In *Procs. International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, pages 1269-1276, Warsaw, 1984. PWN Publ.
- [14] V. I. Oseledec. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19:197-231, 1968.
- [15] Ya. B. Pesin. Families of invariant manifolds corresponding to non-zero characteristic exponents. *Math. USSR. Izv.*, 10:1261-1302, 1976.
- [16] Ya. B. Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Math. Surveys*, 324:55-114, 1977.
- [17] M. Viana. *Lectures on Lyapunov exponents*. Cambridge University Press, 2014.

IMPA, EST. D. CASTORINA 110, 22460-320 RIO DE JANEIRO, RJ, BRAZIL  
*E-mail address*: `elaisc@impa.br`, `viana@impa.br`